12

RADC-TR-81-53
Final Technical Report
May 1981

WIDEBAND SPEECH ENHANCEMENT ADDITION

Queens College

Mark R. Weiss
Ernest Aschkenasy

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, New York 13441

This report has been reviewed by the RADC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RADC-TR-81-53 has been reviewed and is approved for publication.

APPROVED: Edward J Cupples

EDWARD J. CUPPLES
Project Engineer

APPROVED:

JOHN N. ENTZMINGER
Technical Director
Intelligence and Reconnaissance Division

FOR THE COMMANDER: John P. Huss

JOHN P. HUSS
Acting Chief, Plans Office

If your address has changed or if you wish to be removed from the RADC mailing list, or if the addressee is no longer employed by your organization, please notify RADC (IRAA) Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return this copy. Retain or destroy.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19 REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 18 RADC-TR-81-53	2. GOVT ACCESSION NO. AD-A100 462
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 WIDEBAND SPEECH ENHANCEMENT ADDITION,	[redacted] 9 Final Technical Report, Jan[redacted]78—Jul[redacted]80,
	[redacted] N/A
7. AUTHOR(s) 10 Mark R./Weiss Ernest/Aschkenasy	8. CONTRACT OR GRANT NUMBER(s) 15 F30602-78-C-0063 new
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Computer Science Queens College Flushing NY 11367	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 31033G 16 20330303 17 03
11. CONTROLLING OFFICE NAME AND ADDRESS Rome Air Development Center (IRAA) Griffiss AFB NY 13441	12. REPORT DATE 11 May 1981
	13. NUMBER OF PAGES 46
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Same 12 52	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

RADC Project Engineer: Edward J. Cupples (IRAA)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Speech enhancement	Interference/noise reduction
Narrow band noise suppression	
Impulse noise suppression	
Wideband noise suppression	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes the completion of the development and construction of a speech enhancement unit (SEU). This is an electronic instrument that automatically detects and attenuates tones, clicks, and wideband random noises that may accompany speech signals that are transmitted, recorded, or reproduced electronically. An earlier version of this device was tested extensively by the Air Force and then was returned to Queens College for modification and completion of the system. During

cont.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

4 994

cont

the period that is covered by this report, a number of major changes were made in the SEU, leading to a device that is simpler to use, more effective, and more broadly useful in its intended area of application.

Some of the changes that were made in the SEU were aimed at reducing the degree of operator intervention that was required. To this end, the SEU was greatly simplified and made more automatic. The manual Digital Spectrum Shaping (DSS) system and most of the manual controls were removed. A new system was added for adjusting the level of the input signal. It keeps the signal at the level that maximizes the effectiveness of the noise attenuation processes.

The INTEL process for attenuating wideband random noise was incorporated into the SEU. To make it possible for the speech enhancement unit to operate in real-time with all processes active, the hardware and software of the system were modified extensively. The MAP was upgraded by adding a second arithmetic processor and a high speed memory. The existing programs and algorithms were rewritten to reduce their execution times. These and the INTEL programs were modified to fully exploit the capabilities of the upgraded MAP.

CONTENTS

ILLUSTRATIONS

1.0 INTRODUCTION

This report describes the development, testing, and modification of the SEU, an electronic instrument that can be used to enhance the signal-to-noise ratio of speech signals that are accompanied by noise. This speech enhancement instrument is fully automatic and operates in real-time. The signal processing techniques that are implemented in the SEU, as well as the instrument itself, were developed over a number of years, primarily under the sponsorship of the United States Air Force, Rome Air Development Center (RADC). The device that is described in this report was designed and constructed in two stages. The first stage system included processes for automatically detecting and attenuating two types of commonly encountered interference -- tones and impulses. After extensive field tests of the first stage system, work on the SEU was continued with the implemtation of a process for automatic suppression of wideband random noise. The evolution of these processes and the implementation of the processes in the first stage system is described in this section of the report. Section 2 discusses the changes that were made in the device as a result of field tests that were conducted by the Air Force under practical conditions. The modifications that were made to incorporate the process for attenuating wideband random noise are described in Section 3. Finally, in Section 4, recommendations are presented for further development of the system and for a more efficient implementation of it.

1.1 Evolution of the Speech Enhancement Processes

The earliest predecessor of the Speech Enhancement Unit was a crude device that was used in 1967 to demonstrate the feasibility of real-time processing of signals in the spectrum domain. It was built around a real-time spectrum analyzer that was known as a Coherent Memory Filter. Unlike the FFT method of spectrum analysis, which it predated by several years, this instrument was analog in its implementation. It generated both the amplitude spectrum of input signals (in a band from 50 Hz to 1000 Hz), and, almost as a byproduct, an analog version of the complex spectrum of these signals. The potential of this device for processing signals became apparent when it was shown that by use of suitable circuitry, the original signal could be regenerated from the complex spectrum signal. The processing system consisted of four electronic gates. These attenuated arbitrarily set zones in the complex spectrum. When a gate was located such that it attenuated the spectrum region that corresponded to the frequency of a tone that was present in the input to the device, the tone disappeared in the regenerated output signal.

The next evolutionary step was to implement the technique described above in the form of a device that could be used in a practical manner to process speech sounds that had been obscured by tonal noises. This device, which was called a Coherent Spectrum Shaper, incorporated ten electronic gates for the attenuation of tones. The gates were individually adjustable by use of manual controls that set their locations and widths. To aid the user in setting the gates, the system included a display of the spectrum of the input signal. By inspecting the spectrum while listening to the signal he could identify the components of the tones and so locate the gates correctly in the spectrum. A fully automatic version of this process was developed and implemented in the first stage version of the SEU, which consisted primarily of a very powerful minicomputer. The tone attenuation process was implemented as a computer program that generated the spectrum, examined it to identify the components of tones, set gates to eliminate these components from the spectrum, and then regenerated the output signal from the modified spectrum.

Because the process is entirely digital as implemented, it is referred to as digital spectrum shaping, or DSS.

The Coherent Spectrum Shaper also included a method for attenuating wideband random noise. To process this type of noise, the complex spectrum signal was compared to a pair of adjustable thresholds that were symmetrical about a level of zero volts. Spectrum components of the input signal that were lower than the thresholds were set to zero. This type of "baseline" clipping of the spectrum was effective in reducing the level of wideband random noise when the signal-to-noise ratio was greater than 12 dB, but it was ineffective when the S/N was lower than 12 dB.

The failure of the above approach to attenuate wideband random noise effectively led to further research in this area. First, a form of comb filtering was tried, with bandpass filters (actually, electronic gates) set to pass spectrum components at the frequencies of harmonics of the vocal pitch. To determine and track these harmonics accurately, a form of cepstrum pitch extraction was implemented. Although this technique clearly increased the S/N it did not improve the intelligibility or even the listenability of speech signals that were accompanied by random noise. This was because the comb filters passed any signals that were within their bands, and so in spectrum regions outside the formant areas, the filters passed noise. The result was that output speech signals were accompanied by a distracting, buzzy sound whose pitch tracked that of the talker's voice.

The next approach tried was designed to eliminate the undesired noise in the output of the comb filter process. In this method, which was called total spectrum shaping, the amplitude spectrums of input signals that contained speech plus noise were replaced by estimates of the spectrums of the speech component alone. These synthetic amplitude spectrums were then converted into complex spectrums, which were used to generate the output signal. In each synthesized amplitude spectrum, the frequency interval between successive harmonic peaks was made equal to the estimated pitch of the speech in the input signal. The amplitude of each pitch harmonic was made equal to the estimated amplitude of the spectrum peak at the corresponding frequency in the spectrum of the input signal. The shapes of the peaks corresponded to the theoretical shape that would be observed for a constant frequency, constant amplitude signal whose spectrum was determined over the same analysis interval as was the spectrum of the input signal. To obtain a refined estimate of the strength of the speech component at a given pitch harmonic frequency, a form of averaging across three successive spectra was used to form the value of the amplitude of that component in the synthetic spectrum. Finally, a new complex spectrum was generated in which the phase angle at each frequency was the same as that in the original input spectrum, and the absolute amplitude was the same as that in the synthetic spectrum. Then, an output signal was generated by an inverse transformation of the new complex spectrum. This method resulted in a substantial enhancement of the S/N and, in general, avoided introducing spurious noises into the output signal. However, the regenerated speech sounded artificial and, probably due to the averaging process, had a reverberant quality to it.

The technique described above was unsuccessful in itself. However, it led to an important observation: although the influence of noise, in randomizing phase angles, was still present in the synthesized complex spectrum, the absence of noise components in the synthesized amplitude spectrum led to a greatly reduced noise level in the regenerated output signal. This result suggested that an amplitude spectrum in which the amplitudes and shapes of the harmonics corresponded to those that would naturally be observed for a noise-free speech signal could be used to generate such a signal even if the phases of the spectrum components were randomized. To test this concept, amplitude spectrums of noise-free speech signals were obtained and were used to generate complex spectrums in which the phase angles were randomized. Although the resulting output signals sounded "fuzzy", they did not sound either noisy or artificial, and they were fully intelligible.

Clearly, a technique for transforming an amplitude spectrum of speech and noise into the spectrum of only the speech component of the signal would provide the basis for an effective method of enhancing speech that was accompanied by wideband random noise. To develop such a technique, an approach was taken that was radically different from those that had been attempted previously. Underlying it was the concept, similar to that underlying homomorphic filtering, of finding a transformation or a series of transformations that would maximally separate speech from noise. The most effective transformation was found to be the spectrum of the square-root of the amplitude spectrum of the speech and noise. While this function is not the same as the cepstrum, which is the spectrum of the log amplitude spectrum, because it does resemble the cepstrum, and for convenience, in this report it is referred to as the root-cepstrum.

The enhancement procedure consisted of processing the root-cepstrum of an input signal in such a way as to emphasize the components of speech over those of accompanying wideband noise. The enhanced root-cepstrum was then retransformed to generate an enhanced root-spectrum. This was then squared to form an amplitude spectrum with an enhanced ratio of speech-to-noise power.

Because random noise concentrates disproportionately more power in the root-cepstrum region below 0.6 msec than does speech, most of the enhancement that was achieved resulted from the processing of components of the root-cepstrum in this region. The significant difference between the enhancement techniques that were tested lies in the method that was used to process the root-cepstrum components.

The important difference between the root-cepstrum of noise and that of speech is in the distribution of power. For both types of signals, most of the power is concentrated in the root-cepstrum region between zero and 0.5 msec. However, random noise concentrates disproportionately more power in this region than does speech. In fact, for equal amplitude signals, the low-quefrency region of the root-cepstrum will contain about 10 dB more power for noise than it will for speech. All of the schemes for enhancing the S/N of speech in random noise by processing the root-cepstrum operate primarily by attenuating the components in this region. In the region above 0.5 msec, the power in the root-cepstrum of noise falls off monotonically with increasing time (or, equivalently, with increasing quefrency). The root-cepstrum of speech also tends to diminish with increasing time. However, in voiced speech, minor peaks occur at locations that correspond to integer multiples of the pitch period. Broad minor peaks also occur, at time points that correspond to the reciprocal of the frequency interval between formants. For example, the root-cepstrum of the vowel sound /i/, with formants at say 250 Hz, 2250 Hz, and 2750 Hz, will exhibit a broad peak at about 0.5 msec, and one at 2.0 msec.

One method for processing signal components in the root-cepstrum is to set to zero all components in the region from 0.1 msec through 0.5 msec. The component at zero time also can be attenuated, but it must not be eliminated since it represents the DC level of the rooted spectrum. This method (which was given the acronym ASPEN for automatic speech enhancement) was implemented and tested in 1971. It was effective in increasing the S/N, usually by about 6 dB. However, it generated an audible distortion in the regenerated and enhanced speech sounds, giving them an unnatural vocoder-like quality. It also altered the character of the noise, converting the smooth, constant hiss into a sound with a gurgling quality. Some listeners found this to be a more distracting sound than the hiss of the input noise.

Our second approach to processing the signals in the root-cepstrum proved to be more successful. In this method, the average root-cepstrum of the noise in the input signal is, in effect, subtracted from the root-cepstrums of the combined speech and noise. This is the so-called INTEL process. By carefully adjusting the amount of the average root-cepstrum of noise that is subtracted from the root-cepstrums of speech plus noise, the S/N can be enhanced by as much as 12 to 14 dB. The distortions associated with the ASPEN procedure are still present in the INTEL output, but they are very greatly reduced. The INTEL process is one of the three speech enhancement techniques that is implemented in the SEU.

The third of the enhancement processes that was implemented was developed in 1976, during the construction of the first version of the SEU. This process, which is called IMP, is designed to remove impulse noises (e.g., static discharges, ignition noises, etc.) from the input signal. Unlike INTEL and DSS, which operate on transforms of the signal, IMP operates directly on the input signal. As implemented in the speech enhancement unit, IMP is the first of the processes to be applied to the SEU input. In operation, it first examines the time waveform to detect impulses. It next sets to zero regions in the waveform where impulses were found. Finally, IMP fills the resulting gaps with short segments of the signal waveforms adjacent to the gaps. The output of the IMP process is perceived as being continous and free of loud impulses. However, low amplitude thumps can be heard occasionally, usually at points where relatively wide impulses were removed.

The preceding descriptions of the speech enhancement processes were necessarily brief in this capsule history of their development. A more complete description of the INTEL process is presented in section 3, together with a discussion of the problems of implementing it in the SEU. Complete explanations of the DSS and IMP processes can be found in the final technical report on the preceding phase of this project.

1.2 Configuration Development Specification

The speech enhancement unit was developed to provide a means of testing the speech processing techniques under practical conditions. These include the use of the system in a normal working environment to enhance speech signals that are typical of those that are received in the field. Obviously, it was not necessary to develop and implement a production prototype version of the system to meet the stated objective. On the other hand, the device would have to be reasonably practical as regards cost and size. To provide for the possible expansion of the SEU to incorporate other processing techniques, the system had to be made flexible and easy to modify. Finally, and most important, it had to perform the signal processing operations with maximum effectiveness and in real-time.

The processes that are implemented in the SEU perform three types of operations on input signals. These are (1) transformation of the signal to and from the frequency domain, (2) logical tests on and examinations of the signal in either the time or frequency domain, and (3) modification of the signal in either domain. Equipment that performs such operations can be implemented as hardwired circuits, or by a computer that is appropriately programmed, or by a hybrid of these two apprcaches. For reasons of economy, flexibility, and effectiveness, the approach taken was based wholly on the use of a computer. Hardwired circuits would have operated more rapidly and so would have allowed a greater number of real-time processes to be added to the system at a later time. But this approach would have resulted in a system that was far less flexible, far more costly to implement, and probably much larger than the one that was developed.

The basic configuration of the SEU is illustrated in figure 1. Incoming signals are converted in the pre-processor from analog to digital form for input to the computer. At the output of the computer the stream of digital data that represent the processed version of the input signal are converted by the post-processor back to analog form for reproduction as sounds or for recording. These two computer peripherals were designed and constructed at Queens College. The computer itself is a commercially available unit.

Selection of the computer was the single most important decision in the design and development of the SEU. After comparing several units, the Macro-Arithmetic Processor (or MAP) that is manufactured by CSP, Inc., was chosen for use in the SEU. The major factors that led to the selection of the MAP were (1) the speed of the computer, (2) the ease of programming it, (3)

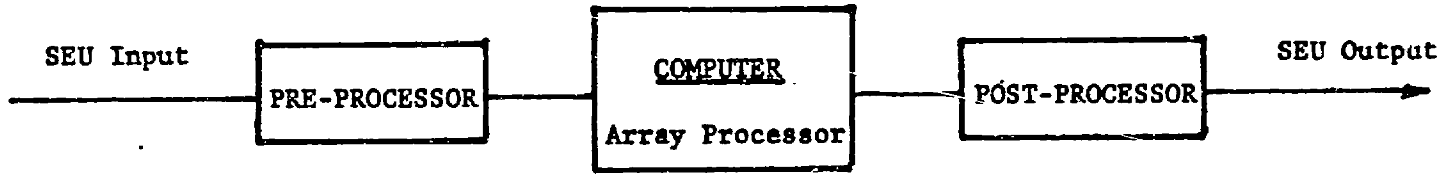

FIGURE 1 BASIC CONFIGURATION OF THE SEU

its precision, (4) its cost, and (5) its size.

The original specifications for the SEU required that it have a dynamic range of at least 60 dB. In binary terms this corresponds to a range of 10 bits. However, such a specification when applied to speech actually requires a sampling range of 14 bits to allow for a (formant 1)/(formant 2) amplitude ratio of up to 24 dB. Processing of the input signals necessarily reduces the total power in the speech components at the same time that the noise components are greatly attenuated. Consequently, the computer had to have a precision greater than 14 bits if any quantization noise or distortion introduced by processing of the signal was to be negligible. Most of the minicomputers and array processors that were available at that time provided at least 16-bit fixed point computation, and so could have provided sufficient precision for the processing of signals with a moderate to high S/N at the input to the SEU. However, it was necessary to provide for the processing of signals in which the S/N was very low. In the case of tonal interference, the objective was to be able to extract speech signals that were as much as 20 dB below the noise level. Hence, a computation range of at least 18 bits was required, plus an additional bit to allow for rounding errors. The processing of speech signals that were obscured by wideband random noise required a similar precision, since at S/N at or below 0 dB the power in the speech signal that is extracted by the INTEL process can be as much as 20 dB below its power in the input.

The second major determinant in the selection of the computer was the required speed of computation. Real-time processing of the signal requires that all of the operations that are performed on a segment of the signal be completed before the succeeding segment is available for processing. That is, the rate of computation must be at least equal to the rate of input of samples of the signal. Since the signal is sampled at a 10-kHz rate, this implies that all calculations per sample must be completed within 100 usec. Actually, only half this ampunt of time is available, since, due to the use of a 50-percent overlapped processing window, each sample is processed twice.

The computation speed that is needed to achieve real-time operation of the SEU was estimated from an analysis of the operations that are performed on the input signal. The most time consuming of these are the transformations between the time and frequency domains. Two such transformations are performed per point by the DSS process and four more are performed by the INTEL process. If, as seemed and was proved reasonable, it is assumed that half the available computation time is consumed by the signal transformations, then each transform must be completed at a rate of 4.16 usec per sample point. This corresponds to a transform computation time of 4.26 msec per 1024 real points.

With the minimum performance requirements of the computer defined, it was possible to identify the device that was most suitable for use in the SEU. All of the standard, commercially available minicomputers were eliminated from consideration because they could not provide the required computation speed. The only way to achieve this speed was to use an array processor. Array processors are special purpose computers that are designed primarily to act as peripheral "number crunchers" to a host computer, receiving data to be calculated on and returning the results when the calculations are completed. After reviewing the characteristics of the array processors that were available, we selected the Macro-Arithmetic Processor (or MAP) for use in the SEU.

The MAP is a dual processor computer. One processor is a very high speed 32-bit floating point arithmetic unit that operates in a pipeline mode. Two such units can be incorporated into a MAP to achieve maximum computation speed. The other processor is a fixed point unit that is used primarily for logical tests on data. Each processor can address any one of three independent memories. Through careful programming, a MAP with two arithmetic processors and at least one high-speed memory can achieve the necessary computation speed. The MAP can be used as a stand-alone data processor, independent of a host computer. Thus, once the SEU software is loaded into the MAP, the time-function samples can be transferred directly between the MAP and the pre- and post-processors. This capability is essential for successful real-time operation of the SEU. Our analyses showed that the time that would have been taken for transfer of data to and from a host minicomputer, even a very high-speed one, would have been too great for real-time operation to be achieved. The other array processors that were available at the time this choice was made were inferior to the MAP in regard to precision or speed, and lacked stand-alone capability. The array processor that was closest in performance to the MAP was less flexible, larger, and about 50 percent more expensive.

The other two system units, the pre-processor and the post-processor, were constructed using standard off-the-shelf components. To meet the accuracy required for the sampling of input signals, the A/D and D/A converters in these units are 14-bit devices. Preceding the A/D converter in the pre-processor is a sample-and-hold unit whose aperature is about 3 nsec. This narrow sampling window permits 14-bit accuracy to be achieved at sampling rates up to 20,000 samples per second, which is twice the rate that is used in the SEU. Preceding the sample-and-hold unit and succeeding the D/A converter are 3.5-kHz low pass filters that prevent any aliasing of components either in the sampled input signal or in the regenerated output signal. Full descriptions of the circuits and characteristics of the pre-processor and post-processor are presenTed in the final technical report for the first stage version of the SEU. 1he changes that were made in the pre-processor circuits are described in section 2 of this report, which follows.

2.0 SIMPLIFICATION OF THE SEU

In December 1977 the U.S. Air Force employed the SEU in a series of field tests to evaluate its usefulness and effectiveness. At that time, the device contained two automatic speech enhancement processes and, to supplement these, it also contained one enhancement process that was manually controllable. To enable the operator to optimize the operation of the SEU, means were provided for adjusting basic system parameters over a wide range. The results of the field tests showed that the automatic enhancement processes were effective in improving the listenability of received speech transmissions and, for signals that were obscured by tones, they improved the intelligibility as well. The tests also revealed that operators of the SEU seldom used either the manual process or the system parameter controls, and that when they were used they frequently were set incorrectly. After reviewing these results, the sponsor decided to eliminate the manual system and the user selectable control of the system parameters and, wherever possible, to automate other controls that previously had to be set manually. These changes are described in this section of the report. Although the changes reduced the flexibility and potential usefulness of the SEU, they also simplified its operation and led to an improvement in the effectiveness of the automatic processes. As a result of removing circuits that were no longer needed, these changes also led to an improvement in the roliability of the system.

2.1 Removal of the Manual DSS System

The manual DSS system in effect permitted the user to selectively attenuate regions of the signal spectrum that contained the components of tones and narrow band noises that failed to be detected and attenuated by the automatic DSS process. To aid the user in locating these regions, a display that showed the spectrum of the signal and the locations of manually set attenuation zones in the spectrum was presented on a CRT. The operator set the locations and widths of the attenuation zones by use of several controls that were located on the panel of the system control unit (SCU). Contained within this unit were the circuits that were needed to convert the display data from digital form, as provided by the MAP, to analog form for input to a display oscilloscope. Inside the MAP, an I/O scroll (which is an interface circuit for peripheral devices) controlled the transfer of data between the SCU and the MAP. The I/O scroll provided the required "handshaking" control signals as well as the necessary strobing of data to and from the I/O bus of the MAP. (A similar I/O scroll was used to control the transfer of speech signal data between the SCU and the MAP.)

Through the use of switches on the SCU panel, the operator of the SEU could also control the duration and the updating of the input data within each input analysis window. He could set the duration of the window (which was referred to as the input signal process-period) to 50 msec, 100 msec, or 200 msec. This feature permitted him to select the process-period that best matched the spectral dynamics of tonal noises in the input signal. He could also "freeze" the input, holding the last process-period length segment of input signal in the MAP memory. This feature permitted the operator to capture transient noises and to locate them in the spectrum. Unfortunately, as in the case of the manual DSS process, these features were seldom used to advantage. Consequently, to further simplify the system, they also were removed.

With the circuits and controls needed by the manual DSS system removed from the SCU, the I/O scroll that handled the signals that were generated by or required by these circuits was left with only three data items to transfer to and from the MAP. These were the three binary signals that turned the automatic processes on or off. To further simplify the SEU, the system was rearranged so that these data are now transferred to the MAP through the same I/O scroll that transfers the input and output signal data. To accomplish this, the process selection signals are multiplexed with the digitized speech signals. As before, the input signal data are read into and out of the MAP at a 10-kHz rate and are stored in the input and output buffers in blocks of 1024 samples. The settings of the process-selection switches are now strobed into the input data stream between the last sample in each group of 128 samples and the first sample in the succeeding group, as shown in figure 2. This change permitted us to remove the I/O scroll that previously controlled the transfer of SCU control data to the MAP. It also permitted us to simplify the cabling between the SCU and the MAP. Previously, it was necessary to use a "patch board" to provide the correct cabling between the four output connectors on the SCU and the edge connectors on the I/O scroll boards. With the removal of the manual DSS system, only two connectors are needed for SCU inputs and outputs. These are now cabled directly to the remaining I/O scroll by use of a fifty-wire ribbon cable.

2.2 ADDITION OF AN AUTOMATIC INPUT GAIN CONTROL

The SEU was designed to process speech signals that are transmitted over standard 3-kHz wide communications channels. The input conditioning circuit

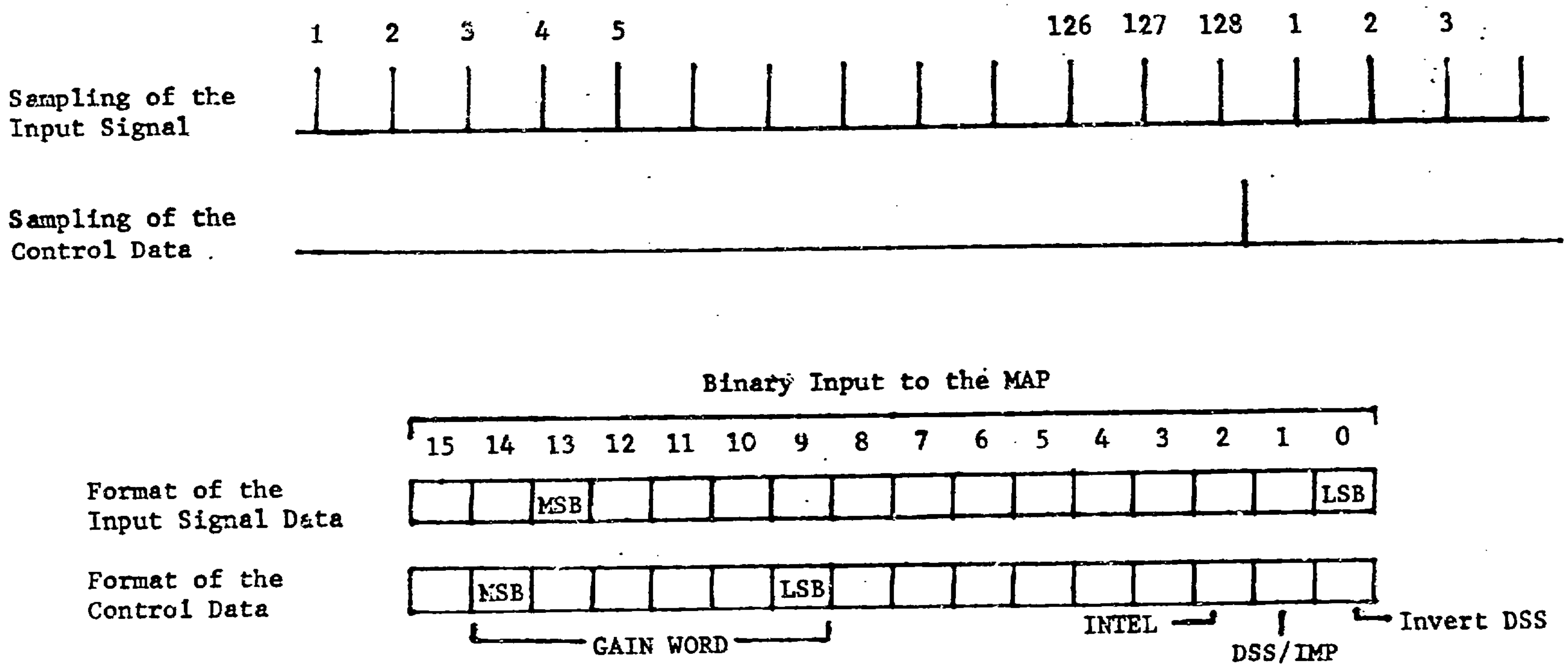
1
2
3
4
5
126
127
128
1
2
3
Sampling of the Input Signal
Sampling of the Control Data
Binary Input to the MAP
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
Format of the Input Signal Data
MSB
LSB
Format of the Control Data
MSB
LSB
GAIN WORD
INTEL
DSS/IMP
Invert DSS

FIGURE 2 MULTIPLEXING OF SEU CONTROL DATA

in the system control unit converts these input signals to digital form for transmission to the MAP, where they are processed. The major electronic components in this circuit are an anti-aliasing low-pass filter, a sample-and-hold, and an analog-to-digital converter. The filter provides over 60 dB of attenuation at frequencies above 5 kHz, and over 80 dB above 8 kHz. For the expected class of input signals, this insures that any aliased components in the input to the MAP will be well below the low end of the 60-dB dynamic range of the SEU. Every 100 microseconds, the sample-and-hold unit samples the input, and "holds" the sampled value until the next sample is taken. The "held" samples are converted to 14-bit binary coded numbers by the A/D converter, at a scale of 14 bits equal to +5 volts and 0 bits equal to -5 volts. The 3-nsec aperature of the sample-and-hold is more than adequately short to insure that for signal frequencies at least up to 10 kHz the held samples reflect the amplitudes of the signals at the sampling instants with an accuracy of at least one part in sixteen thousand (i.e., 1 bit out of 14).

The fourteen-bit range of the A/D converter (ADC) is most effectively used when the peak voltages in the input to this unit approach 5 volts. Since the level of the input signal can vary over a wide range, some means must be provided to adjust this level so as to achieve optimum use of the ADC. In the earlier version of the SEU, this was achieved by preceding the input to the anti-aliasing filter with a high-gain amplifier and controlling the level of the input to this amplifier with a manual gain control (i.e., a potentiometer). However, as was indicated earlier, during the Air Force testing of the SEU it was observed that the operators failed to use this control effectively or to adjust it appropriately when the level of the input signal changed. Therefore, it was decided to supplement this control with an automatic means of optimizing the level of the input signal. This new level control system compresses signals at the input to the sample-and-hold in a 30-dB range from 70 mv to 2 volts to a 6-dB range from 1.25 volts to 2.5 volts. To allow for the possibility that input signals greater than 2 volts may be applied to the SEU, the manual control of the input signal level was not removed from the system control unit. Instead, it was made available to the operator through a selection switch on the panel of the control unit.

The input automatic gain control (AGC) is illustrated in block diagram form in figure 3. The SEU input signal is transmitted to the input conditioning circuit through a multiplying digital-to-analog converter (MDAC). The gain of this device is controlled by a 6-bit binary word, which permits the MDAC gain to be varied over a range of 32 to 1. During each 12.8-msec period, the MDAC gain is held constant, at a level that was set at the end of the preceding period. During each such period, the AGC circuit examines the peak level of the signal at the input to the S/H. From this examination it determines the gain of the MDAC for the succeeding period, with the objective of keeping the signal peaks as high as possible. To allow for rapid increases in the signal level, the AGC does not attempt to force signal peaks to the maximum level. Instead, the circuit adjusts the signal level so that at the input to the sample-and-hold the peak level is in the range 1.25 volts to 2.5 volts, i.e., between 12 dB and 6 dB below the maximum allowable input level to the ADC. It accomplishes this by comparing the full-wave rectified input signal to thresholds that are set at these levels. If during any 12.8-msec segment of the input signal the signal level exceeds the upper threshold, the gain of the MDAC is reduced by 6 dB during the next segment. On the other hand, if the signal level falls below the lower threshold, the gain is

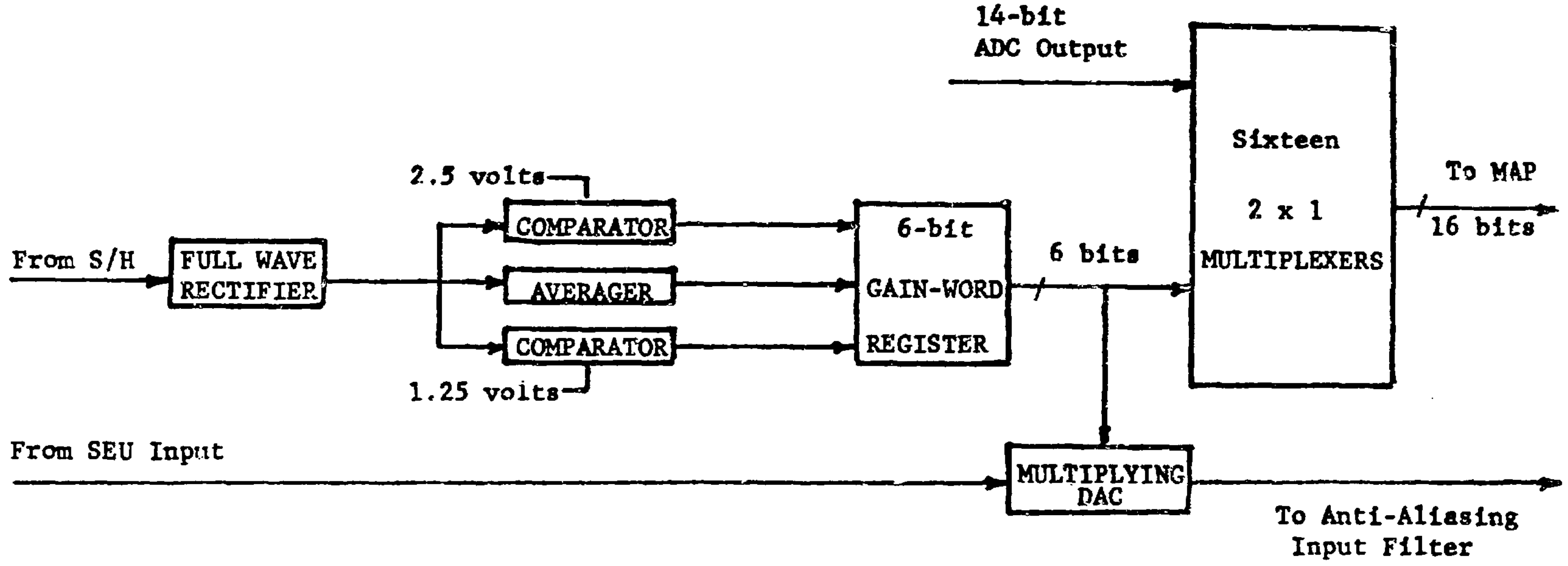

FIGURE 3 BLOCK DIAGRAM OF THE INPUT AGC

increased by 6 dB during the succeeding segment. If the maximum level falls between the thresholds, the gain is left unchanged. The 12.8-msec segment length was chosen to insure that at least one pitch peak would be compared to the thresholds during each segement of the signal.

By making the peak signal amplitudes in each 12.8-msec segment fall within a narrow range, the AGC distorts the relative amplitudes of successive segments. This distortion must be removed before the signal can be processed. To accomplish this, at the end of each 12.8-msec period the binary word that set the MDAC gain is transmitted to the MAP. In the MAP, the samples within each group of 128 samples are rescaled by dividing them by the numerical value of their associated gain word.

The AGC approach outlined above has two potential weaknesses. If an input signal contains speech that is accompanied by very large impulses (e.g., static or ignition noise) the gain of the AGC will be reduced in successive 6 dB steps until the impulse peaks are in the 1.25 volt to 2.5 volt range. This can result in excessive reduction in the level of the underlying speech. For example, if the impulse peaks are 20 dB above the speech peaks (which would make them about 30 dB above the average speech level) the AGC would reduce the speech level to 36 dB below the peak input level to the A/D converter. Consequently, the speech signal would be sampled by only 8 bits out of the 14-bit range of the converter, and the regenerated signal at the output of the SEU would exhibit sampling quantization noise. To avoid this possibility and assure the maximum effective use of the sampling range of the A/D converter, the reduction of the gain of the AGC is halted when the average level of the full-wave rectified input signal falls below a minimum level. The second potential weakness is that the input signal can be clipped if it increases by more than 6 dB in a 25.6-msec interval and definitely will be clipped when it increases by more than 12 dB in this interval. However, the latter circumstance is unlikely to occur for signals of the type that the SEU was intended to process. These are likely to be accompanied by a significant level of noise and may have been subjected to some degree of dynamic compression by the transmitting or receiving equipment in the communication channel. In either case, the effect will [illegible] limit the maximum relative change in signal level that can be observed in a 25.6-msec interval and so reduce the chance that peaks in the speech signal will be clipped.

The circuit diagram of the AGC unit is shown in figure 4. The input signal is applied to pin 17 of IC7a, a Datel multiplying D/A converter. The output of the MDAC is taken through a unity gain operational amplifier to the AGC selector switch on the front panel of the system control unit. The input signal also is applied to the input of a unity gain inverting amplifier (IC13a) whose output is rectified and summed with the rectified input signal by another section of this amplifier (IC13b). The full wave rectified signal is fed to a solid state switch (IC15a) and to a pair of comparators (IC14a and IC14d). The output of either comparator will drop to a low level (i.e., to a logical 0) when the rectified signal exceeds the constant voltage that is applied to its alternate input. These outputs drive the "set" inputs of a pair of flip- flops (IC1a and IC1b). When IC1a is set its output is high (logical 1). When IC1b is set its output is low. Both flip-flops are reset by a pulse (GRSL) that is applied to their "clear" inputs.

The gain of the MDAC is controlled by a 6-bit binary word that is stored in IC6, an 8-bit shift register. When the system is turned on and the I/O scroll begins to run, it develops a level signal (RSH). The rising edge of

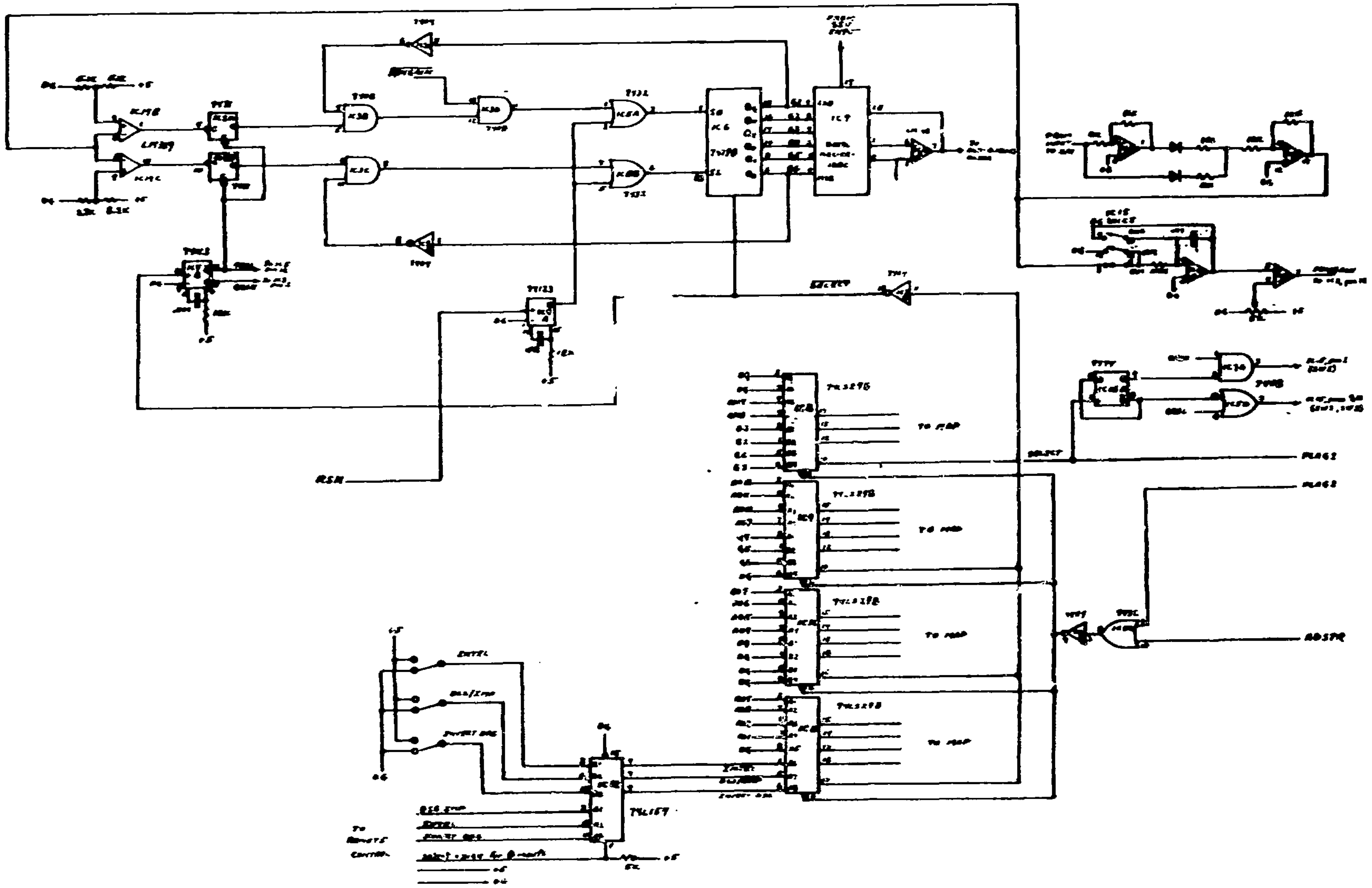

FIGURE 4 CIRCUIT DIAGRAM OF THE INPUT AGC

RSH triggers a monostable multivibrator, IC4a, which produces a 1-usec wide pulse. This pulse, applied through a pair of OR gates (IC5a and IC5b) to inputs S0 and S1 of IC6 cause a preset binary word to be loaded into the register. In binary form this word is 00001000. The middle six bits of the word are applied to the gain control inputs of the MDAC. When the gain transfer trigger, GXFRH, is applied to pin 11 of the register, the stored word will shift to the right, or left, or not at all depending on the status of the inputs S0 and S1. The word will shift to the left if S1 is high and S0 low, and to the right if the reverse is true. It will not shift at all if both inputs are low. The upper three bits of the gain word are also applied to three of the four "B" inputs of a quad 2x1 multiplexer (IC8). The lower three bits are applied to the B inputs on a similar unit (IC9). The upper six bits of the digitally converted samples of the input signal are applied to the A inputs on these multiplexers and the lower six bits to A inputs on two other multiplexers, IC10 and IC11. IC11 also receives the binary signals that enable or disable the operation of the automatic enhancement processes. The four A inputs on each of these devices are selected for transfer to the device outputs when the signal at pin 10 is a logical 0. When this signal is a logical 1 the B inputs are selected for transfer to the output. The multiplexer outputs are transmitted to the MAP.

The sequence of operations in the AGC is illustrated by the waveforms shown in figure 5. Every 100 usec, the input conditioning circuit generates a pulse (ADSTR) that strobes the A/D converter data from the A inputs of the quad multiplexers into the registers in the multiplexer outputs. ADSTR is applied through an OR gate (IC5c) and an inverter (IC2b) to pin 11 of each multiplexer. For 128 of these pulses, the multiplexer input at pin 10 is at a

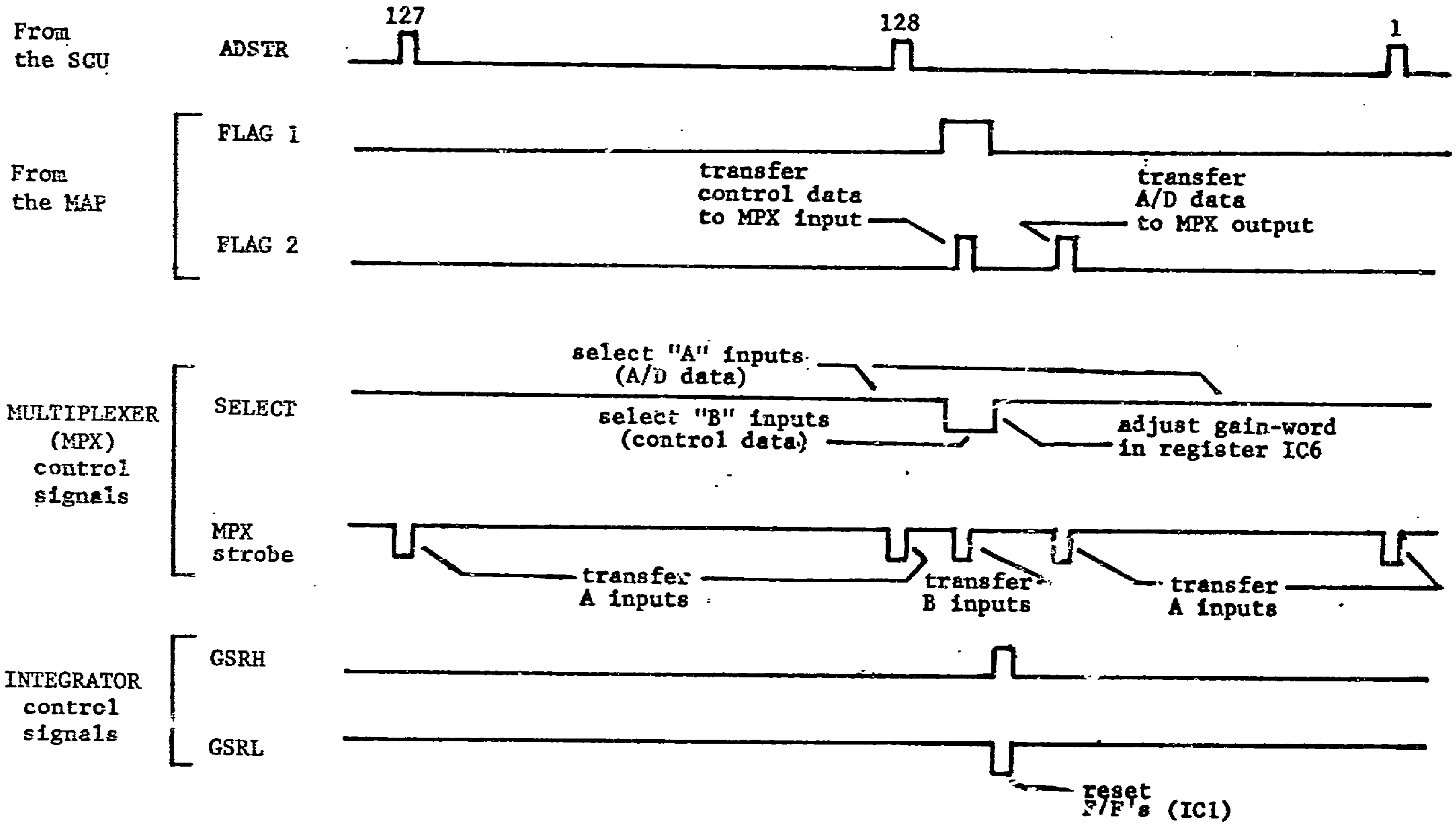
From the SCU
ADSTR
127
128
1
From the MAP
FLAG 1
FLAG 2
transfer control data to MPX input
transfer A/D data to MPX output
MULTIPLEXER (MPX) control signals
SELECT
select "A" inputs (A/D data)
select "B" inputs (control data)
adjust gain-word in register IC6
MPX strobe
transfer A inputs
transfer B inputs
transfer A inputs
INTEGRATOR control signals
GSRH
GSRL
reset F/F's (IC1)

FIGURE 5 AGC TIMING DIAGRAM

logical 0 level, and so ADSTR will strobe the A input data into the multiplexer output register. After the 128th pulse, the I/O scroll in the MAP generates a pulse signal, FLAG1, that raises the multiplexer select line to a logical 1 and so causes the B inputs to be selected for transfer to the multiplexer outputs. At the same time, FLAG1 is inverted by IC2e and applied to pin 11 of the gain data register, IC6, and to a monostable multivibrator, IC4b. However, since both of these IC's trigger on a 0 to 1 transition, the inverted leading edge of FLAG1 will not cause a change in their outputs.

Shortly after the scroll generates the leading edge of FLAG1 it produces a second, shorter pulse, FLAG2. This signal, applied like ADSTR, strobes the selected input data into the output registers of the multiplexers. These data consist of the process-selection binary signals and the gain control word that established the gain of the MDAC during the preceeding 12.8 msec. Within 1.5 usec, these data are transferred via the I/O scroll to the MAP. Subsequently, FLAG1 is permitted to drop to a zero level. The inverted trailing edge of FLAG1 triggers IC4b which resets flip-flops IC1a and IC1b. However, before the resetting of these IC's is felt at inputs S0 and S1 of IC6, the inverted trailing edge causes the gain word currently stored in IC6 to shift right, left, or not at all, depenuing on the signal levels at S0 and S1 at that instant.

During the next 12.8 msec, the newly established gain word will control the gain of the MDAC. The full-wave rectified input signal will be compared to thresholds of 1.25 volts and 2.5 volts by comparators IC4a and IC4d. If, during this period, the signal never exceeds either threshold, the output levels of flip-flops IC1a and IC1b will remain in the reset state (i.e., IC1a low and IC1b high). Therefore, when FLAG2 is generated the S0 input of IC6 will be low and S1 will be high. Consequently, the gain word will shift one bit to the left, which will increase the gain of the MDAC by 6-dB during the succeeding 12.8 msec. If the input signal exceeds the lower threshold but not the upper one, IC1b will be set, causing its output to drop to a logical 0 level, and the signals at both S0 and S1 will be low. As a result, the gain word will not shift and so the AGC gain for the next 12.8-msec interval will be the same as it was for the current one. Finally, if both thresholds are exceeded, IC1a also will be set and so its output will be high while that of IC1b will be low. This will result in a one-bit right shift of the gain word, producing a 6-dB decrease in the MDAC gain when FLAG2 is generated.

It is necessary to prevent the gain word from shifting entirely out of the register when either no signal or a continuous high-level signal is present. This is accomplished by transmitting the flip-flop outputs through AND gates IC3b and IC3c. The alternate inputs to those gates are complemented versions of the levels at those outputs of IC6 that correspond to the extreme right and extreme left shift positions of the 1 bit in the gain word. Thus, when the word is 100000 (maximum gain) the level at pin 10 of IC3c will be low, which will prevent any further leftward shift of the word. Similarly, when the AGC gain is a minimum, the gain word will be 000001, and so the inverted output of pin 18 on IC6 will produce a 0 level input at pin 4 of IC3b.

Earlier in this section of the report the need to prevent unnecessa y excesive reduction in the level of input speech signals was discussed. The technique that was described for accomplishing this is implemented by applying the gated output of flip-flop IC1a to the S0 input of IC6 through a second AND gate, IC3d. The alternate input to this gate is a level signal that is high when the average level of the full-wave rectified input signal is above a minimum acceptible level, and low otherwise. The level signal, MINGAIN, is generated by an integrator circuit (IC13c) and a comparator (IC14b). Solid state switches S1, S2, and S3 (IC's 15a, 15b, and 15c) clear the integrator every 25.6 msec. This length of integration period was selected to insure that the computed average signal level would include at least two pitch periods of an input speech signal. The switches are controlled by waveforms that are generated by flip-flop IC16b, AND gate IC3a, and OR gate IC5d. The outputs of the flip-flop are complemented at every occurrence of FLAG1. When the flip-flop is set (that is, when Q is high and Q' is low) the output of IC3a will be low and that of IC5d will be high. Consequently, SW1 will be closed, connecting the integrator input to the output of IC13b, and the short circuit across the integrator capacitor (SW3) will be open as will be SW2. When the next FLAG1 is generated, the flip-flop outputs will be complemented. Until GRSH is generated the SW1 control signals will be the same as before. So will be the SW2 and SW3 control signals, since GRSL will be high. At the end of FLAG1, GRSH will go high and GRSL will go low. Consequently, for the duration of this 1.5-usec long pulse, SW2 will connect the input of the integrator to ground and SW3 will short the integrator capacitor, discharging it.

3.0 IMPLEMENTATION OF INTEL IN THE SEU

Implementing INTEL as a real-time process and integrating it with the DSS and IMP processes raised a number of major design problems. These can be grouped into four categories:

1. Whether to execute INTEL before or after DSS and IMP.
2. How to provide the memory space that would be required to store the programs and data arrays of INTEL.
3. How to achieve real-time operation of the SEU with all processes active.
4. How to generate the cepstrum threshold function for live input signals in which the wideband noise component can vary.

Most of the time and effort expended on this project was devoted to working out the solutions to these problems, as is described in this section of the report.

3.1 Location of INTEL in the Sequence of Processes

The sequence in which signal processing operations are performed can greatly affect the effectiveness of the processes. For example, during the implementation of the earlier version of the SEU it was found to be better to perform IMP after DSS rather than before it. Executing IMP first did not significantly improve the ability of DSS to detect tones since the spectrum levels of the components of impulses tend to be small even when compared with those of tonal noises. On the other hand, removing strong tones from the time waveform of the signals before it is processed to remove impulses significantly increased the likelihood of detecting weak impulses. Similar observations and considerations led us to conclude that it would be better to place INTEL after DSS and IMP rather than before them. Placing INTEL first might have removed sufficient wideband noise from input signals to enable DSS to detect and attenuate weak tones. However, if INTEL had been placed first, any strong tones in the input signal would have distorted the shape of the root-cepstrum, particularly in the low-quefrency region. Since the multipliers that scale the cepstrum threshold in this region were selected to achieve the optimum attenuation of a pure random-noise root-cepstrum, any distortion of the shape of the root-cepstrum due to the presence of tones would have reduced the effectiveness of INTEL. To avoid any loss in the effectiveness of INTEL, it was made the last of the processes to be executed in the SEU.

3.2 Implementation Problems

One of the two major problems of implementing INTEL and integrating it into the SEU was related to providing the memory space that was needed to store the INTEL programs and arrays in the MAP. At the start of this project, the programs and arrys of DSS and IMP required virtually all of the 12,288 word storage capacity of the MAP. Although the INTEL software had not yet been written, it was possible to estimte that the programs and arrays would require at least 1500 words in the MAP's memory. While some of the needed space could have been (and ultimately was) provided by rearranging data arrays in DSS and IMP and by equivalencing arrays wherever possible, these measures would not have been adequate and the size of the MAP memory had to be expanded.

The second, and more difficult problem was that of minimizing the execution time of the signal processing techniques so as to achieve real-time operation of the SEU. This required that all operations be completed by the time each new segment of input signal is ready for processing. That is, all operations had to be completed within 102.4 msec, since the input signal, which is sampled at a 10-kHz rate, is stored in the SEU in blocks of 1024 samples. At the start of this project, DSS and IMP took about 90 msec to run, leaving about 12 msec of idle time. This was grossly inadequate for the requirements of INTEL. The FFT's of INTEL alone would have required 80 msec if they were computed at the same speed as those of DSS. Obviously, it was essential to greatly reduce the time taken for these calculations and 'or all of the other signal processes that were implemented in the SEU.

A three step approach was used to achieve real-time operation of the SEU. First, the MAP was modified so as to increase the speed at which it can perform calculations. Second, the FFT programs were rewritten so as to take maximum advantage of the improved capability of the MAP, and thereby cut their running time by 50 percent. Third, the existing DSS and IMP routines were revised so as to speed them wherever possible, and they were rewritten so as to make full use of the MAP hardware, as was the newly written INTEL programs. The need for the last of these steps can be demonstrated by considering the time required for the execution of these routines. With the time required for the calculation of an FFT cut in half, the time needed for the DSS FFT's became about 20 msec and the time needed for the INTEL FFT's became 40 msec. This left 42.4 msec for the completion of all of the processing operations. The DSS and IMP operations took 50 msec in the original version of the SEU. Even if the time required by these operations could also be cut in half (which it could not) there would be only 17.4 msec left for the INTEL processes. These processes include amplitude weighting of the time waveform, detection of voicing, root-compression of the spectrum, computation of an averaged and scaled root-cepstrum of noise, and subtraction of cepstrum threshold from the root-cepstrum of the incoming signals. Therefore, to maximize the time that was available for the INTEL processes, the existing DSS and IMP programs had to be changed to reduce their execution times to the absolute minimum.

It was apparent at the outset that there was another way to increase the time available for INTEL, one that would have required much less effort. This would have been to increase the duration of each 1024-point sample block of input data and thereby extend the time that was available for the completion of all processes. For example, the block duration could have been increased from 102.4 msec to 128 msec by reducing the sampling rate to 8 kHz. The additional 25.6 msec might have made it possible to avoid the need to modify some if not all of the existing processing programs. However, the cost in system performance would have been too great. With the number of signal points processed per block held constant, a 20-percent reduction in the sampling rate would have resulted in a 20-percent reduction in the band-range of the processed spectrum. In particular, the DSS spectrum range would had to have been reduced to 2400 Hz. This would have resulted in a substantial loss in the quality of the regenerated speech sounds. To avoid any degradation in the high quality of the DSS output, this approach was not considered appropriate at the start of work on this project. However, it was not rejected completely, but was held in reserve against the possibility that the steps that were to be taken to speed the SEU might be insufficient to achieve real-time operation.

3.3 Upgrading the MAP

Substantial changes were made in the MAP to provide space for the programs and arrays of INTEL and to speed the calculations and logical operations that were required in all of the processes. The MAP is a very powerful minicomputer. Its architecture provides for three completely independent data busses, with a separate memory associated with each bus. Programs can be executed by a conventional central processing unit (called the CSPU in the MAP). Very high-speed calculations on arrays of data can be performed by a separate processor, called the AP or arithmetic processor. Both the CSPU and the AP can communicate with any of the three memories and they can operate in parallel. Thus, the AP can be used to perform array calculations at the same time that the CSPU can be performing logical operations.

At the start of this project, 4096-word memories were provided on all three busses of the MAP. All three memories were composed of MOS integrated circuits and each had a nominal access time of 500 nsec. To help meet the space and speed requirements of the final version of the SEU, the 4K memory on BUS1, in which the programs and some data arrays are stored, was replaced by an 8K memory. The 2K memory on Bus 3, in which the FFT data arrays are stored, was replaced by a 3K, 180-nsec bipolar memory. The faster memory more than halves the time required to fetch and store FFT data. The addition of 1K of storage to Bus 3 memory permitted the cosine table that is needed for the calculation of the FFT's to be stored in the same high-speed memory as the FFT data.

A second major change that was made in the MAP was to add a second arithmetic processor. Two AP's, operating in parallel, can be used to reduce the time that is needed to perform calculations on an array of data. This speedup is especially effective for algorithms, such as the FFT, in which the same calculations are performed on the entire array. In such cases, with careful programming of the parallel operation of the AP's, the FFT calculation time can be reduced by nearly 50 percent. (Although the use of two AP's will double the rate at which calculations can be performed, the time it takes to compute the FFT of an array cannot be cut in half because of time losses that result from contention for the same memory by both AP's.)

The upgrading of the array processor, now classified as a Model 300 MAP, was completed early in the project. To provide the DC power that was needed by the new components of the MAP (the second arithmetic processor and the 3K bipolar memory) the MAP power supply was replaced by a more powerful one that provides an additional 22 amperes at 5 volts DC.

3.4 Speedup of the FFT Computations

In the earlier version of the SEU, Fourier transforms were computed by a radix-2, not-in-place algorithm. A 2048-point real-to-complex or complex-to-real transform was calculated in 20 msec. The first step toward speeding up this calculation was to substitute a radix-4 FFT algorithm for the one in use. Then, after verifying that the algorithm was computing the transforms correctly, it was reprogrammed to use both AP's. Lastly, the cosine table was moved from Bus 1 memory, where it had been stored together with the FFT program, to the bipolar memory on Bus 3. These changes reduced the time to compute 2048-point transforms to 9.8 msec. The DSS and INTEL processes require together six transforms of 2048 points. Four of these are time/spectrum conversions, each of which transform 2048 points during the processing of each 102.4 msec long block of input signal data. The remaining transforms are the four 512-point spectrum-to-cepstrum transforms and the four 512-point cepstrum-to-spectrum transforms that are performed during the INTEL processing of each block of input data. Thus, all of the FFT calculations together take 58.8 msec, which leaves 43.6 msec available for the DSS, IMP, and INTEL processing routines.

3.5 Speedup of DSS and IMP

In the earlier version of the SEU it took about 50 msec to execute the DSS and IMP processes. Most of this time was consumed by routines, such as MATCHPK, that ran in the CSPU. This is the routine that identifies as components of tones those spectrum peaks that appear on two successive spectra and that have the same amplitudes and locations on both spectra. This routine, and others like it, require computer operations that examine the data and/or compare them to a reference, and, based on the result, take appropriate actions. These same operations, if run in the AP, could be executed in about one-tenth the time that they take to run in the CSPU. Unfortunately, the AP has very limited capability for performing logical tests and for making decisions, and none whatever for program branching . However, through careful use of the AP instruction repetoire, it was possible to program some of these routines to run in the AP. It also was possible to convert some of the DSS and IMP array calculation programs from MAP 200 form (i.e., computations that use a single arithmetic processor) to MAP 300 form that uses two arithmetic processors operating in parallel. hese changes reduced the execution times of DSS and IMP to about 20 msec.

3.6 Generation of the Cepstrum Threshold

The cepstrum threshold is a computed function that is subtracted from the root-cepstrum of the input signal in order to enhance the ratio of signal power to noise power in the root-cepstrum. It is generated by computing the average root-cepstrum of the noise in the input signal and then multiplying each of three regions of that function by an appropriate scale factor. These regions are the quefrency range from 0.1 msec through 0.5 msec, the range above 0.5 msec, and the zero time point. The scale factors that are used were determined by experiment as being those that produce the optimum cepstrum threshold function. The optimum function is the one that maximizes the attenuation of noise while keeping the distortion of speech to a minimum.

Ideally, the average root-cepstrum from which the cepstrum threshold is generated should continuously reflect the average properties of the noise that is present in the input to the SEU. To permit this function to respond to changes in the quality and intensity of the input noise, the average root-cepstrum must be updated continuously. Moreover, to insure that the average root-cepstrum represents only the noise in the input signal, the individual root-cepstrums that contribute to the average must represent segments of the input signal that contain no speech sounds. The first of these requirements was met by using a lossy integration function to compute the average root-cepstrum. To meet the second requirement, an algorithm for detecting voiced speech sounds was made part of the INTEL software. This algorithm cannot detect unvoiced speech sounds and it is not a perfect detector of voiced ones (especially at S/N below 0 dB). However, it is sufficiently sensitive to detect voiced speech sounds which, if they were included in the calculation of the average root-cepstrum of noise, would significantly distort its shape.

3.6.1 Calculation of the Average Root-Cepstrum of Noise

A number of methods can be used to compute a running average of a function. The simplest is to use lossy integration, i.e., to allow the contribution of a component to diminish increasingly with time. This is inplemented in the SEU by the formula current avg. root-cepstrum = (1 - a) (previous avg. root-cepstrum)

+ (a) (current root-cepstrum)

The value of the constant, a, was chosen to yield a time-constant of 0.5 seconds, thereby permitting the calculated average root-cepstrum to track reasonably rapid changes in the input noise. The average of the root-cepstrum is updated only when the voicing detection algorithm indicates that voiced speech is not present in the input signal. At all other times the average of the root-cepstrum is held constant.

3.6.2 Detection of Voiced Speech Sounds

The voicing detection procedure that was implemented in the SEU is a version of the NUPITCH technique for measuring vocal pitch. This technique, which was developed under previous Air Force contracts, is capable of detecting voiced speech sounds at S/N down to -6 dB. AT 0 dB, the type 1 error (false rejection of voicing) is about 5 percent and the type 2 error (false detection of voicing) is about 15 percent. At -6 dB, the type 1 error rate increases to about 10 percent. Most of the type 1 errors occur at instants when the instantaneous intensity of voiced speech sounds is well below the average intensity of the speech sounds (e.g., at the starts and ends of sentences, at some syllable boundaries, just before vowel/fricative transitions, etc).

NUPITCH detects voicing by determining if the largest spectrum peaks in the spectrum in the region from 200 Hz to 1000 Hz are harmonics of some fundamental frequency. The routine first identifies the five largest peaks in this range and then selects the three largest of these and arranges them in order of increasing frequency. The frequency intervals between the first and second peaks and between the second and third peaks are determined and compared in a series of tests in which the larger of the two intervals is compared first to the smaller one, then if necessary, to twice the smaller one and finally to three times the smaller one. If, for any of these comparisons the compared frequency intervals differ by less than 40 Hz, the spectrum peaks are said to be conditionally harmonically related. If none of the comparisons satisfies this requirement, the smallest of the three peaks is dropped, and the entire procedure is repeated with the fourth largest peak substituted for the third one and, if necessary, once again with the fifth largest peak substituted for the third one. If, after all peaks have been used, none of the comparisons satisfies the spacing requirement, the decision is made that the input signal did not contain voiced speech sounds.

When three conditionally harmonic peaks are found, the algorithm tests their harmonicity more precisely. First, the difference in frequency between the first and last peaks in the sequence is divided by the total harmonic interval between them. The resulting number is an estimate of the fundamental frequency of the presumably harmonically related peaks. Next, the frequency of each peak is divided by the estimated fundamental frequency and the result is rounded to the nearest integer multiple of 0.5. If any one of these calculations leads to a non-integer result, the pitch interval is cut in half and the calculated values are doubled. These are the probable harmonic numbers of the spectrum peaks. Next, the estimated fundamental frequency is multiplied by each probable harmonic number and the results subtracted from the frequency of the corresponding spectrum peak. Finally, the average of the absolute differences, determined as above, is computed. If this value is less than 10 Hz, the input signal is identified as containing voiced speech whose pitch is as computed. Otherwise, the algorithm decides that the input signal did not contain voiced speech.

Because of the relatively wide limit imposed on the average absolute difference measure, the NUPITCH procedure tends to detect voicing in speech when it is not present far more often than it fails to detect voicing when it is present. For the purposes of INTEL, this distribution of errors is desirable, since it tends to insure that any sound that even remotely resembles voiced speech will not be included in the calculation of the average root-cepstrum of noise.

Most of the time that is taken by the NUPITCH routine is consumed during the initial step of identifying peaks in the amplitude spectrum. To minimize the running time of the routine, the peak-detection procedure, which could most easily have been programmed to run in the CSPU, was programmed to run in the AP.

4.0 OBSERVATIONS AND RECOMMENDATIONS

The speech enhancement unit as implemented meets or exceeds all of its required performance characteristics. In its present form, it can be used under a wide range of practical conditions to test all of the processing techniques, without requiring any special training of the operators of the system. For the most part, the operator's role is reduced to turning the automatic processes on or off, when and as required. The SEU can be used as a stand-alone device or, through the use of the remote control feature, as a part of a larger speech processing station.

The last few weeks of work on the SEU were devoted to testing, adjusting, and "fine tuning" the system. The results of the tests are summarized in the observations that are described below and in the recommendations that follow.

4.1 Observations

4.1.1 INTEL

The operation of the INTEL process is controlled by four parameters. These are: the duration of the process period, the location of the boundary between the low and high quefrency band in the root-cepstrum, the root that is used for compression of the spectrum, and the factors that are used to scale the average root-cepstrum of noise. In the current version of the SEU, only the last of these parameters can be adjusted. The others are fixed at values that, during the early years of research on the INTEL process, were found to be optimum for processing speech sounds that were accompanied by white noise.

The cepstrum scale factors critically affect the performance of INTEL.

If they are set too high, the normal dynamic variations of sound intensity in speech will be greatly exaggerated in the INTEL output. The gurgling quality that is associated with the noise in the INTEL output also will be exaggerated. If the scale factors are set too low, the INTEL process will not provide as great an enhancenent of S/N as it is capable of doing. Selection of the optimum scale factors proved to be crucial to achieving the best possible performance of INTEL.

The scale factors that were selected were determined heuristically. A series of tests were conducted during which the scale factors were adjusted for optimum processing of each of a number of test recordings. These included recordings of typical communications signals that were provided by RADC, several recordings that were provided by federal and local investigative agencies, and a recording that was made in the cockpit of a commercial aircraft shortly before it crashed. The background noises ranged from white to pink, and the estimated S/N in the test signals ranged from 10 dB to -6 dB. The optimum scale factors for the zero quefrency component of the root-cepstrum tended to cluster at a value of 0.75. For the low-quefrency band (0.1 msec through 0.5 msec) the optimum scale factor was 0.5, and for the band above 0.5 msec it was 0.5.

It was observed that the optimum values did not change with changes in the spectral distribution of the noise. Thus, unlike some enhancement procedures that process the signal in the spectrum domain, it is not necessary in INTEL to keep track of the distribution of wideband random noise.

Informal listening tests showed that the gurgling quality in the regenerated noise at the output of the INTEL process was greatly reduced over that in previous versions of the process. At the same time, the enhancement of S/N was far greater, ranging between 12 dB and 14 dB. The improved performance of INTEL, as implemented in the SEU, over previous simulated versions of the process is due entirely to the use of optimum cepstrum function scale factors. On the average, the intelligibility of the regenerated speech signals was neither improved nor reduced when compared with that of the input signals. On the other hand, the greatly reduced noise level in the output, together with the reduced gurgling quality in the noise, led to a significant improvement in the listenability of the speech in the INTEL output.

4.1.2 DSS and IMP

Although neither the DSS nor the IMP process was modified in the final version of the SEU, it was observed that their performance was improved over that in the first stage version of the device. We believe that this improvement is a result of the inclusion of an automatic input gain control in the pre-processor. This device continuously maximizes the input signal level and thereby assures maximum use of the 14-bit dynamic range of the anaolg-to-digital conversion system. (In effect, the AGC adds five more bits to this range.)

4.2 Recommendations

Based on our observations of and our experience with the SEU we can recommend a number of changes that would make the system more useful and more effective. Some of these could be implemented quickly and economically in the current version of the SEU. Others would require some additional research or development.

4.2.1 Changes Not Requiring Research

The INTEL process as designed and implemented works very well for signals in which the backgr^ound noise level either is constant or changes at some reasonably continuous rate. However, in practice it sometimes happens that the SEU input signal disappears for a brief period and then reappears. This will occur, for example, when an FM transmission drops below the squelch level of a receiver and then rises above it. In the present system, the cepstrum threshold function diminishes steadily during the time that the signal is lost. At the same time, the output AGC gain increases steadily. Consequently, when the signal is restored, the audio output of the SEU is initially very loud. Then, as the proper cepstrum threshold and AGC levels are regenerted, the output level returns to the same value it had before the signal disappeared. These initial bursts of sound level can be very annoying and can affect the intelligibility of the speech sounds that immediately follow them. The bursts can be eliminated entirely by changing the INTEL learning process and the output AGC process so that, at the option of the operator, the AGC gain level and the cepstrum threshold are held constant from the moment he activates appropriate switches. During the time that these functions are held constant, the AGC gain level and the average root-cepstrum of the input noise would be updated continuously in the normal manner. Thus, as soon as the operator releases the "holds" on the cepstrum threshold and on the AGC gain level, correct values for these functions would be available immediately.

A second recommended change is to restore some manual gate capability to the DSS process. It sometimes occurs that when wideband random noise that accompanies a signal has been attenuated by INTEL, a few tones that were obscured by the noise are revealed. Even when tones are loud enough to be audible in the input signal, the presence of random noise can make it impossible for the automatic DSS process to detect them. The availability of manual gates would permit the operator of the SEU to attenuate such tones when they occur.

A third recommended change is to restore some ability to change the duration of the DSS process period. The 200-msec period that is immplemented in the current version of the SEU is satisfactory for stable or slowly changing tones, but it is too long for the DSS process to detect rapidly changing tones. The addition of a capability for changing the process period to 100 msec when desired would extend the usefulness of the SEU.

4.2.2 Changes Requiring Additional Research or Development

The INTEL process attempts to distinguish between the root-cepstrums of speech and of random noise. The ability to discriminate between these classes of signals is maximized when the analysis window is optimally matched to the spectral dynamics of the components of speech throughout the spectrum. The use of single analysis window for all spectrum components, as in the current version of INTEL, makes it necessary to set the window length to a compromise value. By experiment, this value was found to be about 50 msec. Ideally, a 100-msec window would best match the speech components below 750 Hz, a 50-msec window would be best for the range 750 Hz to 1500 Hz, and a 25-msec window for the range above 1500 Hz. The use of a variable window should be explored to evaluate what advantages, if any, it would provide over the use of a single, fixed window.

A second change that should be examined is the substitution of a trapezoidal function to weight the input signal for the triangular function that currently is used. The current function requires a 50 percent overlap of successive analysis windows. Consequently, each sample of the input signal must be processed twice. With a trapezoidal weighting function, the overlap could be reduced substantially, with a consequent reduction in the number of sample points that must be processed per second. The result would be an increase in the time that is available for other processes that could be implemented in the SEU.

The final recommended change is to use an EPROM memory to permanently store the SEU software inside the MAP. With the programs instantly available, the SEU could automatically be ready for use in less than one-tenth of a second after power was turned on or after it was restored if a power interruption occurred. With such a memory in place, the SEU would be free of the need to be connected to an external device to load the SEU programs into the MAP, and the unit would become a self-contained, mobile device.